Intelligent Agent for Science Data Processing

Keith Golden* Ramakrishna Nemani Wanlin Pang† Petr Votava‡
*Google Inc.
NASA Ames Research Center
†University of California, Santa Cruz
‡California State University, Monterey Bay
wpang@email.arc.nasa.gov

Petabytes of remote sensing data are now available from Earth-observing satellites to help measure, understand and forecast changes in the Earth system, but using these data effectively can be surprisingly hard. The volume and variety of data files and formats are daunting. Simple data management activities, such as locating and transferring files, changing file formats, gridding point data, and scaling and reprojecting gridded data, can consume far more personnel time and resources than the actual data analysis. We address this problem by developing a planner-based agent for data production, called IMAGEbot, that takes data product requests as high-level goals and executes the commands needed to produce the requested data products.

IMAGEbot is based on automated constraint-based planning and a flexible component-based architecture. Unlike more traditional approaches, where the instruction sequences for managing and processing data are hand-coded; in our agent-based approach, the instruction sequences are automatically generated based on user requests and available data sources. New data sources, models or data-processing programs can be added in a plug-and-play fashion, and the planner can adapt to errors or data dropouts by trying alternative ways of achieving the same goal, such as using other, possibly lesser quality, data sources.

We have demonstrated this technology in the Terrestrial Observation and Prediction System (TOPS), an ecological forecasting system that assimilates data from Earth-orbiting satellites and ground weather stations to model and forecast conditions on the surface, such as soil moisture, vegetation growth and plant stress. The planner identifies the appropriate input files and sequences of operations needed to satisfy a data request, executes those operations on a remote TOPS server, and displays the results, quickly and reliably.

Whereas TOPS is concerned with geospatial data measuring specific variables of the Earth system, such as precipitation, vegetation productivity and fire risk, the agent-based approach is applicable to other types of data as well. Our current effort focuses on extending IMAGEbot functionalities and applying it to other areas.